Plan Overview

A Data Management Plan created using DMPonline

Title: Head and neck paraganglioma registry

Creator: Carolijn de Bresser

Principal Investigator: Dr. J.A. Rijken, Dr. B.J. Petri

Data Manager: Dax Steins

Affiliation: UMC Utrecht

Funder: UMC Utrecht

Template: UMC Utrecht DMP

Project abstract:

Head and neck paraganglioma are rare, mostly benign neoplasmata, developing out of the ectodermal neural crest. These slow growing tumors may become malignent, with a 5-year survival of 50%. In order to conduct more impactful research, the UMC Utrecht's department of Vascular surgery aims to setup a registry to collect high quality long term follow-up data for future research.

ID: 85128

Start date: 01-10-2021

Last modified: 26-05-2023

Copyright information:

Head and neck paraganglioma registry

1. General features

1.1. Please fill in the table below. When not applicable (yet), please fill in N/A.

DMP template version	30 (don't change)
ABR number *(only for human-related research)*	-
METC number *(only for human-related research)*	22-008
DEC number *(only for animal-related research)*	-
Acronym/short study title	HNPGL
Name Research Folder	22-008_HNPGL
Name Division	Surgical Specialties
Name Department	Vascular Surgery
Partner Organization	None
Start date study	01-10-2021
Planned end date study	T.b.d.
Name of datamanager consulted*	D. Steins, P Ojha, N. Koning
Check date by datamanager	07-10-2021, 21-12-2021, 26-5-2023 (amendement)

1.2 Select the specifics that are applicable for your research.

- Use of Questionnaires
- Multicenter study
- Prospective study
- Non-WMO

Registry from the department of Vascular Surgery, UMC Utrecht, the Netherlands. The PI of this study is dr. B.J. Petri (vascular surgeon).

2. Data Collection

2.1 Give a short description of the research data.

Primary Objective: to create a registry that can be used for future research on optimising diagnostic protocols, treatment strategies, improving symptom-free survival and optimising patient follow-up among HNPG patients.
Data flow: After broad consent has been obtained, clinical information from the UMC Utrecht's electronic health records (EHR; HiX) made available via the Research Data Platform (RDP; structured data in MS SQL Server/SAS Enterprise) will be (partially) extracted by the division datamanager. This dataset will be pseudonymized with a key-linking table and stored in secure research folder on the UMCU's network drive. This dataset, including data not available in the RDP, will be collected in Castor EDC (eCRF) by members of the research team. Castor EDC is a browser-based, metadata-driven EDC software solution and workflow methodology for building and managing online clinical research databases. The eCRF contains data items as specified in this research protocol.
Additional radiological scans (i.e. MRI, CT scans) from PACS will be pseudonymized and made available for research through the UMCU's Research Imaging Architecture (RIA).

Subjects	Volume	Data Source	Data Capture Tool	File Type	Format	Storage space
Human	Ongoing	EPD (HiX)	Research Data Platform (RDP)	Quantitative	.csv / /xlsx	0-10GB
Human	Ongoing	eCRF	Castor EDC	Quantitative	.csv .xslx .sav	0-10GB
Human	Ongoing	PACS/US/MRI/CT scanner	Research Imaging Architecture	Imaging	.dcm	101-1000 GB

2.2 Do you reuse existing data?

- Yes, please specify

In this retrospective study, we use data from HiX. For this study, it is not possible to get a pseudonymized dataset (e.g. from the RDP), because preliminary literature review showed that, although a great deal of data exists about the technological aspects of the subject, there is little data available to answer our specific research question. There is thus a need to collect primary data on this topic.

2.3 Describe who will have access to which data during your study.

My division datamanager receives a datamart from theResearch Data Platform (RDP) that contains direct identifying personal data (e.g. date of birth) and pseudonymized data. The datamanager is authorized to link different datasets of the selected patient group and thus has access to personal data such as patientID. The key table linking study specific IDs to patient IDs is available to the datamanager and members of the research team.

Type of data	Who has access
Direct identifying personal data	Research team, Datamanager
Key table linking study specifics IDs to patients IDs	PI, Datamanager, registry coordinator J.M. de Bresser
Pseudonymized date	Research team, datamanager

2.4 Describe how you will take care of good data quality.

Clinical information from patients will be partially collected in an eCRF in Castor--a certified Data Capture Tool. In the eCRF, skips and validation checks are built in. Data quality will be checked by the registry coordinator. Data exports for future studies will be frozen before analysis. Data from the RDP and RIA will be matched by study subject code.

#	Question	Yes	No	N/A
1.	Do you use a certified Data Capture Tool or Electronic Lab Notebook?	X		
2.	Have you built in skips and validation checks?	X		
3.	Do you perform repeated measurements?	X		
4.	Are your devices calibrated?			X
5.	Are your data (partially) checked by others (4 eyes principle)?	X		
6.	Are your data fully up to date?	X		
7.	Do you lock your raw data (frozen dataset)	X		
8.	Do you keep a logging (audit trail) of all changes?	X		
9.	Do you have a policy for handling missing data?	X		
10.	Do you have a policy for handling outliers?	X		

2.5 Specify data management costs and how you plan to cover these costs.

#	Type of costs	Division ("overhead")	Funder	Other (specify)
1.	Time of datamanager	X		
2.	Design of eCRF	X		
3.	Data capture tool license fee	X		
4.	Storage	X		
5.	Archiving	X		
6.	Questionnaire license fee	X		

2.6 State how ownership of the data and intellectual property rights (IPR) to the data will be managed, and which

agreements will be or are made.

In this multicenter registry, all participating centers are owner of the data. As paragangliomas are a relatively rare disease we want to encourage centers to enter their data in the registry and to be involved in the research. All participating centers only have access to their own data, but can get access to pseudonomised data upon request. All this is stated in contracts for research collaboration and a data transfer agreement.

3. Personal data (Data Protection Impact Assessment (DPIA) light)

Will you be using personal data (direct or indirect identifying) from the Electronic Patient Dossier (EPD), DNA, body material, images or any other form of personal data?

- Yes, go to next question

I will process personal data. I have consulted the division management and I do not have to complete a full DPIA. I therefore fill out this DPIA light and proceed to 3.1.

3.1 Describe which personal data you are collecting and why you need them.

Which personal data?	**Why?**
Patient characteristics (year of birth, sex)	To describe our study population
Medical history (history of diseases, family history date start of symptoms, specifics about referral, data send with the referral)	To specify the patients background
Hospital diagnostics (physical exam, lab results, radiology scans/reports, tumor pathogenicity, genetic tumor mutation, type of intervention and its details, complications)	To analyse the different diagnostic tools and treatments.
Follow-up (development of other paragangliomas/tumors or recurrence and its treatment, symptoms of enhanced lab results, treatment of these results, radiology scans, hospital visits, symptoms/physical exam/additional diagnostics and treatment during follow-up, reason lost to follow-up)	Keep monitoring the patient, follow development in diagnostics and therapy.
Questionnaires	To analyse the quality of life of the patient and keep monitoring the patient's quality of life.

3.2 What legal right do you have to process personal data?

- Study-specific informed consent

Broad consent

3.3 Describe how you manage your data to comply to the rights of study participants.

Right of Access	Research data are coded, but can be linked back to personal data, so we can generate a personal record at the moment the person requires that. This needs to be done by an authorized person.
Right of Rectification	The authorized person will give the code for which data have to be rectified.
Right of Objection	We use informed consents.
Right to be Forgotten	In the informed consent we state that the study participant can stop taking part in the research. Removal of collected data from the research database cannot be granted because this would result in a research bias. Patients can choose upon withdrawel if researchers can still use the patient's date or, that the researchers can't use this data anymore, but that the data will still be available.

3.4 Describe the tools and procedures that you use to ensure that only authorized persons have access to personal data.

1. We use the secured Research Folder Structure that ensures that only authorized personnel has access to personal data, including the key table that links personal data to the pseudoID.
2. We make use of a certified Electronic Data Capture (EDC) tool (Castor). To send surveys, email address will be used in the EDC, but this is encrypted for the users in such a way that users can send emails to subjects without seeing the actual email address. No personal data other than email address will be used in the EDC.
3. Patient digital imaging data for study purposes will be stored at the Research Imaging Archive (RIA) facility of the imaging division of UMC Utrecht. For safe processing of images, RIA will be used (uses pseudonymization in order to guarantee safe processing). Only authorized personnel can access the (pseudonymized) imaging in the RIA container via personal login. The linkage table for the pseudonymized images will also be stored at the RIA. The container can only be accessed by users with the proper rights. Hospitals may transfer digital data into the RIA through secure connections. The RIA shields patient identifiable information through pseudonymized identifiers (i.e., study number) and only allows access to authorized researchers.

3.5 Describe how you ensure secure transport of personal data and what contracts are in place for doing that.

1. We will not transport any personal data outside the UMCU network drives.

2. In case we need to transport personal data with colleagues, we use Surffilesender with encryption.

3. We have a Research Agreement and/or Data Transfer Agreement with Erasmus MC, LUMC and Amsterdam UMC. The agreement is stored at location (secured Research Folder Structure of the UMC Utrecht): L:\Onderzoek\Vaatchirurgie\22-008_HNPGL\B_Documentation\6_Contracts

4. Data Storage and Backup

4.1 Describe where you will store your data and documentation during the research.

UMC Utrecht is initiator of this multicenter study. All data and documentation collected by the UMC Utrecht will be stored in the secured Research Folder Structure of the UMC Utrecht. Importantly, personal data is stored separately from other research data and adequate access and control rights are in place. In other participating sites, data and documentation will be stored accordingly. For analysis, data will be stored in DRE.

4.2 Describe your backup strategy or the automated backup strategy of your storage locations.

1. All (research) data is stored on UMC Utrecht networked drives from which backups are made automatically twice a day by the division IT (dIT).
2. During data collection, automatic backups will be made in the Electronic Data Capture Tool Castor. Upon completion of data collection, all data are exported and saved in the Research Folder Structure where they are automatically backed up by the UMC Utrecht backup system.

5. Metadata and Documentation

5.1 Describe the metadata that you will collect and which standards you use.

Patient data will be extracted from HIX in the Research Data Platform (RDP). This metadata will be handled in an Excel-file with a codebook. This will be placed in the secured online research folder by the data manger of the department.
For the data collected in Castor, a codebook of my research database is available in Castor.

5.2 Describe your version control and file naming standards.

We will distinguish versions by indicating the version in the filename of the master copy by adding a code after each edit, for example V1.1 (first number for major versions, last for minor versions). The most recent copy at the master location is always used as the source, and before any editing, this file is saved with the new version code in the filename. The file with the highest code number is the most recent version and older versions are moved to a folder OLD. The major versions will be listed in a version document (projxVersDoc.txt), stating the distinguishing elements per listed version.

6. Data Analysis

6 Describe how you will make the data analysis procedure insightful for peers.

N/A

7. Data Preservation and Archiving

7.1 Describe which data and documents are needed to reproduce your findings.

1. The data package will contain: the raw data, the study protocol describing the methods and materials, a codebook with explanations on the variable names, and a 'read_me.txt' file with an overview of files included and their content and use.

7.2 Describe for how long the data and documents needed for reproducibility will be available.

Data and documentation needed to reproduce findings from this non-WMO study will be stored for at least 15 years.

7.3 Describe which archive or repository (include the link!) you will use for long-term archiving of your data and whether the repository is certified.

After finishing the project, the data package will be stored at the UMC Utrecht Research Folder Structure and is under the responsibility of the Principal Investigator of the research group. When the UMC Utrecht repository is available, the data package will be published here.

7.4 Give the Persistent Identifier (PID) that you will use as a permanent link to your published dataset.

To be determined

8. Data Sharing Statement

8.1 Describe what reuse of your research data you intend or foresee, and what audience will be interested in your data.

1. My peers will be reusing all research data in the final dataset to generate new research questions.
2. The raw data can be of interest for other researchers or for spin off projects.

8.2 Are there any reasons to make part of the data NOT publicly available or to restrict access to the data once made publicly available?

- Yes (please specify)

T.b.d.

8.3 Describe which metadata will be available with the data and what methods or software tools are needed to reuse the data.

T.b.d.

8.4 Describe when and for how long the (meta)data will be available for reuse

- Other (please specify)

To be determined.

8.5 Describe where you will make your data findable and available to others.

T.b.d.